MASSACHUSETTS INSTITUTE OF TECHNOLOGY
Department of Electrical Engineering and Computer Science

6.438 ALGORITHMS FOR INFERENCE
Fall 2014

Quiz 1

Thursday, October 16, 2014
7:00pm–10:00pm

- This is a closed book exam, but two $8\frac{1}{2}'' \times 11''$ sheets of notes (4 sides total) are allowed.
- Calculators are **not** allowed.
- There are **3** problems.
- The problems are not necessarily in order of difficulty. We recommend that you read through all the problems first, then do the problems in whatever order suits you best.
- Record all your solutions on the exam paper. We have left enough space for each part. Extra blank sheets and stapler are available in case you need more space. You may want to first work things through on the scratch paper provided and then neatly transfer to the exam paper the work you would like us to look at. Let us know if you need additional scratch paper.
- A correct answer does not guarantee full credit, and a wrong answer does not guarantee loss of credit. You should clearly but concisely indicate your reasoning and **show all relevant work**. Your grade on each problem will be based on our best assessment of your level of understanding as reflected by what you have written in the answer booklet.
- Please be neat—we can't grade what we can't decipher!

Problem 1

Part 1 - Modelling:
Consider the following events that may happen in a typical student dorm:

A *fire alarm* (A) can be set off due to multiple reasons, upon which the students may have to *evacuate* (E) the building. One possibility for the alarm setting off is that some student accidentally *burns their food* (B) while cooking, which then causes a *fire* (F) in the building. Another possibility is that a student has some *flammable chemicals* (C) in their room, which can accidentally start a fire, causing the fire alarm to trigger. It is also possible that the fire alarm is *defective* (D), and goes off for no reason at all. All of these result in the poor students being evacuated from their rooms. At times, the rooms may get flooded with water because of *broken plumbing* (G), which also requires students to evacuate their rooms.

(a) (**2 Point**) Draw a directed graphical model over the variables A, B, C, D, E, F, G.

Part 2 - Independences:

For this part of the problem, we will consider only *pairwise independences* i.e. independences of the form $X \perp\!\!\!\perp Y|Z$, where X and Y are sets consisting of a single random variable, and Z is an arbitrary set of variables. With respect to the graphical model obtained in part 1, answer the following questions:

(b) (**2 Point**) List all unconditional pairwise independences (i.e. $\mathrm{Z} = \phi$) among the variables A, B, C, D, E, F, G. You should write these independences in a 7x7 table, with a '1' indicating that variables are independent, and '0' indicating they are dependent.

(c) (**2 Point**) Now suppose that you are given the variable A, i.e. you know whether or not the fire alarm has gone off. Re-draw the above table for pairwise independencies conditional on A.

Note: In parts (b) and (c), we only care about independencies of the form $X \perp\!\!\!\perp Y|Z$ where X, Y, and Z are disjoint. You may write an 'x' for the remaining entries.

Part 3 - Probabilities:
Next we will assign potentials to the nodes and use them to calculate the probability of certain events. To do this, we first define the associated RV's explicitly.

Variables: For each event (e.g. A), we assign a binary random variable (e.g. X_A), which can take values 0 or 1, representing whether or not event A occurs.

Potentials: For each node in the DAG, we assign potentials as follows: Assume a node R has k parents. If all of these parents take the value 0, X_R also takes the value 0 with probability 1. Else, use the following rules:

1. For i $\in \{1, 2, \ldots k\}$: $\mathbb{P}_{X_R|X_{\Pi_R}}(1|$ exactly i parents out of k have value 1$) = 1 - \frac{1}{2i}$
2. For each node R that does not have any parents, we assign $\mathbb{P}_{X_R}(1) = \frac{1}{2}$

(d) (**1 Point**) Calculate the probability of burnt cooking given that the fire alarm has gone off i.e. find $\mathbb{P}(X_B = 1|X_A = 1)$.

(e) (**1 Point**) Repeat the above calculation with the additional information that the fire alarm is also defective i.e. $X_D = 1$.

Part 4 - Triangulation:
Recall that a graph is said to be *chordal* if every cycle of length 4 or more in the graph has a 'shortcut' or 'chord', i.e. if some 2 non-consecutive vertices in the cycle are connected by an edge. We call the process of converting a graph into a chordal graph by adding edges as *Triangulation*.

(f) (**2 Point**) Triangulate the following graph, using as few edges as possible:
Hint: Elimination algorithm

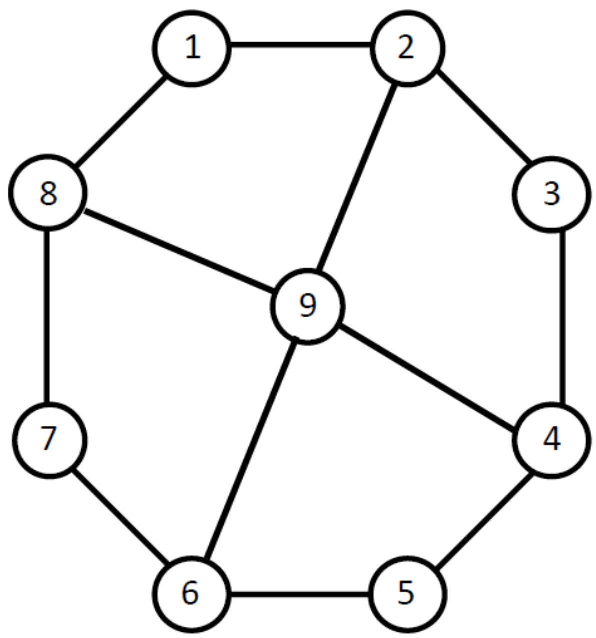

Figure 1: Problem 1(f)

Problem 2

Consider the directed graphical model below, and answer the following questions. Notice Part 1 and Part 2 can be attempted independently.

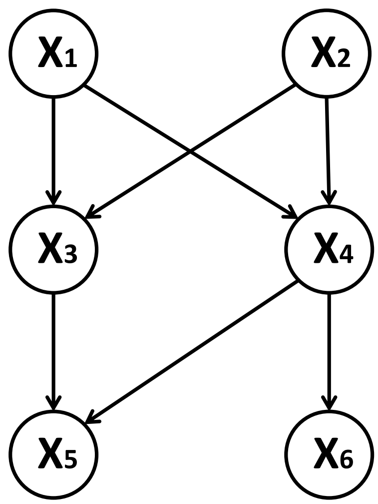

Figure 2: Original Directed Graph

Part 1 - Conversions Among Different Types of Graphical Models

(a) (**1 Point**) Assume the directed graph is a P-map for the associated distribution. Find the minimal undirected I-map for this distribution.

(b) (**1 Point**) Compare the original directed graph and the undirected graph you found in (a), are there any conditional or unconditional independence that is satisfied by the original directed graph but not the undirected graph? List all such independences. If there isn't any, write empty set.

(c) (**1 Point**) Turn the undirected graph you got from (a) into a factor graph.

(d) (**1 Point**) Find a minimal directed I-map for the factor graph you got in (c) using ordering $1, 2, 3, 4, 5, 6$.

Part 2 - Gaussian Graphical Models

For part 2, consider a multi-variate Gaussian distribution over $\mathbf{x}_1$, $\mathbf{x}_2$, $\mathbf{x}_3$, $\mathbf{x}_4$, $\mathbf{x}_5$, $\mathbf{x}_6$ for which Figure 2 is a P-map. Let J be the information matrix for the Gaussian distribution.

(e) (**3 Point**) In the table below, indicate which entries of J must be zero. Place '0's in those entries and leave the rest empty. Provide a brief explanation of your reasoning.

J =

	X1	X2	X3	X4	X5	X6
X1						
X2						
X3						
X4						
X5						
X6						

Figure 3: Problem 2(e)

(f) (**3 Point**) If we marginalize over $\mathbf{x}_3$, we get a new multi-variate Gaussian over variables $\mathbf{x}_1$, $\mathbf{x}_2$, $\mathbf{x}_4$, $\mathbf{x}_5$, $\mathbf{x}_6$. Denote the information matrix of this new distribution as J'. Indicate which entries of J' must be zero. Place '0's in those entries and leave the rest empty. Provide a brief explanation of your reasoning.

$J' =$

	X1	X2	X4	X5	X6
X1					
X2					
X4					
X5					
X6					

Figure 4: Problem 2(f)

Problem 3

This problem explains how graphical model inference can be used for Bayesian inference with mixture distribution model. Such a model captures various scenarios in practice including Natural Language Processing, Social Data Processing, Finance, etc.

Part 1

We start with the simplest possible mixture distribution, a binary valued (i.e. takes value in $\{0, 1\}$) random variable, denoted by X, which is made of mixture of K components. We describe this precisely as follows.

Let there be K different coins, each with different biases, $p_1, \ldots, p_K \in [0, 1]$. That is, when the k^{th} coin is tossed, with probability p_k the outcome is head $(= 1)$ and with probability $1 - p_k$ the outcome is tail $(= 0)$, for $1 \leq k \leq K$. These are K different components.

To generate mixture of the above K components, we utilize a multinomial random variable that takes value in $\{1, \ldots, K\}$. Precisely, let T be a random variable such that $\mathbb{P}(\mathsf{T} = k) = \mu_k$, for $1 \leq k \leq K$. Then the random variable X is generated as follows: first sample T; then toss coin number k if $\mathsf{T} = k$, $1 \leq k \leq K$; and outcome of this coin is equal to X.

(a) (**2 Point**) Write down a graphical model description of the above mixture distribution.

(b) (**2 Point**) Compute $\mathbb{P}(X = 1)$ using the description in (a). Explain how this can be done using sum-produce algorithm.

Part 2

Now we shall consider more interesting setting. Let us described mixture distribution over $N+1$ binary valued random variables, $(\mathsf{X}; \mathsf{Y}_1, \ldots, \mathsf{Y}_N)$, with K mixture components.

As before, let the mixture component be denoted by multinomial random variable T with $\mathbb{P}(\mathsf{T} = k) = \mu_k$, for $1 \le k \le K$. To generate $(\mathsf{X}; \mathsf{Y}_1, \ldots, \mathsf{Y}_N)$, we start by first sampling component T. Then, as before, we generate X by tossing coin of bias p_k if $\mathsf{T} = k$, for $1 \le k \le K$.

Now $(\mathsf{Y}_1, \ldots, \mathsf{Y}_N)$ are generated as follows. We are given K (known) N-dimensional vectors $\Theta^1, \ldots \Theta^K \in \mathbb{R}^N$. If $\mathsf{T} = k$, then we use vector Θ^k and generate $\mathbf{Y} = (\mathsf{Y}_1, \ldots, \mathsf{Y}_N) \in \{0, 1\}^N$ as

$$\mathbb{P}\big(\mathbf{Y} = \mathbf{y}\big) \propto \exp\big(\sum_{i=1}^{N} \theta_i^k y_i\big), \tag{1}$$

for $\mathbf{y} = (y_1, \ldots, y_N) \in \{0, 1\}^N$ (we used notation $\Theta^k = (\theta_1^k, \ldots, \theta_N^k) \in \mathbb{R}^N$.

(c) (**2 Point**) Write down a graphical model description of the above described mixture distribution.

(d) (**2 Point**) Describe sum-product algorithm to compute $\mathbb{P}(X = 1|\mathbf{Y} = \mathbf{y})$ given observation $\mathbf{Y} = \mathbf{y}$, using the description in (c).

(e) (**2 Point**) Describe sum-product algorithm to compute $\mathbb{P}(\mathsf{Y}_1 = 1)$, using the description in (c).